18.05 Lecture 36
May 11, 2005

Review of Test 2 (see solutions for more details)

Problem 1:
$\mathbb{P}(X = 2c) = \frac{1}{2}, \mathbb{P}(X = \frac{1}{2}c) = \frac{1}{2} \rightarrow \mathbb{E}X = 2c(\frac{1}{2}) + \frac{1}{2}c(\frac{1}{2}) = \frac{5}{4}c$
$\mathbb{E}X_n = (\frac{5}{4})^n c$

Problem 2:
$X_1, ..., X_n$
n = 1000
$\mathbb{P}(X_i = 1) = \frac{1}{2}, P(X_i = 0) = \frac{1}{2}$
$\mu = \mathbb{E}X = \frac{1}{2}, \mathrm{Var}(X_1) = p(1-p) = \frac{1}{4}$
$S_n = X_1 + ... + X_n$
$\mathbb{P}(440 \leq S_n \leq k) = 0.5$

$$\frac{S_n - n\mathbb{E}X_1}{\sqrt{n\mathrm{Var}(X_1)}} \rightarrow \frac{S_n - 1000(1/2)}{\sqrt{1000(1/4)}} = \frac{S_n - 500}{\sqrt{250}}$$

$$\mathbb{P}(\frac{440-500}{\sqrt{250}} \leq z \leq \frac{k-500}{\sqrt{250}}) = 0.5$$

by the Central Limit Theorem:

$$\approx \Phi(\frac{k-500}{\sqrt{250}}) - \Phi(\frac{440-500}{\sqrt{250}}) = \Phi(\frac{k-500}{\sqrt{250}}) - \Phi(-3.75) = \Phi(\frac{k-500}{\sqrt{250}}) - 0.0001 = 0.5$$

Therefore:

$$\Phi(\frac{k-500}{\sqrt{250}}) = 0.5001 \rightarrow \frac{k-500}{\sqrt{250}} = 0, k = 500$$

Problem 3:

$$f(x) = \frac{\theta e^\theta}{x^{\theta+1}} I(x \geq e); \psi(\theta) = \frac{\theta^n e^{n\theta}}{(\prod x_i)^{\theta+1}} \rightarrow \max$$

Easier to maximize the log-likelihood:

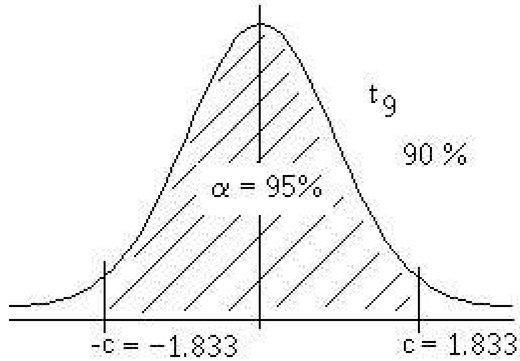
$$\log \psi(\theta) = n\log(\theta) + n\theta - (\theta+1)\log \prod x_i$$

$$\frac{n}{\theta} + n - \log \prod x_i = 0 \rightarrow \theta = \frac{n}{\log \prod x_i - n}$$

Problem 5:
Confidence Intervals, keep in mind the formulas!

$$\overline{x} - c\sqrt{\frac{1}{n-1}(\overline{x^2} - \overline{x}^2)} \leq \mu \leq \overline{x} + c\sqrt{\frac{1}{n-1}(\overline{x^2} - \overline{x}^2)}$$

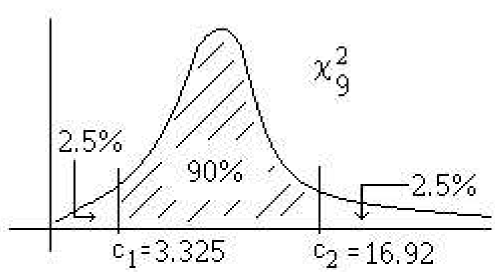
Find c from the T distribution with n - 1 degrees of freedom.



Set up such that the area between -c and c is equal to $1 - \alpha$
In this example, c = 1.833

$$\frac{n(\overline{x^2} - \overline{x}^2)}{c_2} \leq \sigma^2 \leq \frac{n(\overline{x^2} - \overline{x}^2)}{c_1}$$

Find c from the chi-square distribution with n - 1 degrees of freedom.



Set up such that the area between c_1 and c_2 is equal to $1 - \alpha$
In this example, $c_1 = 3.325, c_2 = 16.92$

Problem 4:
Prior Distribution:

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

$$f(x_1, ..., x_n|\theta) = \frac{\theta^n e^{n\theta}}{(\prod x_i)^{\theta+1}}$$

Posterior Distribution:

$$f(\theta|x_1, ..., x_n) \sim f(\theta) f(x_1, ..., x_n|\theta)$$

$$\sim \theta^{\alpha-1} e^{-\beta\theta} \frac{\theta^n e^{n\theta}}{(\prod x_i)^\theta} = \theta^{\alpha+n-1} e^{-\beta\theta+n\theta} e^{-\theta \log \prod x_i} = \theta^{(\alpha+n)-1} e^{-(\beta-n+\log \prod x_i)\theta}$$

Posterior $= \Gamma(\alpha + n, \beta - n + \log \prod x_i)$
Bayes Estimator:

$$\widehat{\theta} = \frac{\alpha + n}{\beta - n + \log \prod x_i}$$

Final Exam Format
Cumulative, emphasis on after Test 2.
9-10 questions.
Practice Test posted Friday afternoon.
Review Session on Tuesday Night - 5pm, Bring Questions!

Optional PSet:
pg. 548, Problem 3:
Gene has 3 alleles, so there are 6 possible combinations.
$p_1 = \theta_1^2, p_2 = \theta_2^2, p_3 = (1-\theta_1-\theta_2)^2$
$p_4 = 2\theta_1\theta_2, p_5 = 2\theta_1(1-\theta_1-\theta_2), p_6 = 2\theta_2(1-\theta_1-\theta_2)$
Number of categories → r = 6, s = 2.
2 Free Parameters.

$$T = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2_{r-s-1=3}$$

$\psi(\theta_1, \theta_2) = \theta_1^{2N_1}\theta_2^{2N_2}(1-\theta_1-\theta_2)^{2N_3}(2\theta_1\theta_2)^{N_4}(2\theta_1(1-\theta_1-\theta_2))^{N_5}(2\theta_2(1-\theta_1-\theta_2))^{N_6}$
$= 2^{N_4+N_5+N_6}\theta_1^{2N_1+N_4+N_5}\theta_2^{2N_2+N_4+N_6}(1-\theta_1-\theta_2)^{2N_3+N_5+N_6}$

Maximize the log likelihood over the parameters.
$\log\psi = const. + (2N_1+N_4+N_5)\log\theta_1 + (2N_2+N_4+N_6)\log\theta_2 + (2N_3+N_5+N_6)\log(1-\theta_1-\theta_2)$
Max over $\theta_1, \theta_2 \rightarrow$
$\log\psi = a$
$log\theta_1 + b$
$log\theta_2 + c$
$log(1-\theta_1-\theta_2)$

$$\frac{\partial}{\partial\theta_1} = \frac{a}{\theta_1} - \frac{c}{1-\theta_1-\theta_2} = 0; \frac{\partial}{\partial\theta_2} = \frac{b}{\theta_2} - \frac{c}{1-\theta_1-\theta_2} = 0$$

Solve for θ_1, θ_2

$$\frac{a}{\theta_1} = \frac{b}{\theta_2} \rightarrow a\theta_2 = b\theta_1$$

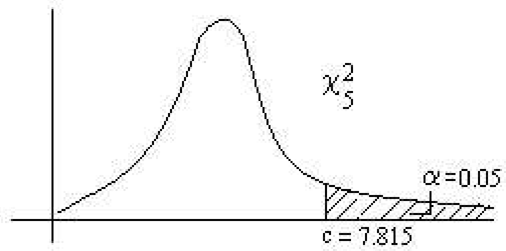
$a - a\theta_1 - a\theta_2 - c\theta_1 = 0, a - a\theta_1 - b\theta_1 - c\theta_1 = 0 \rightarrow$

$$\theta_1 = \frac{a}{a+b+c}, \theta_2 = \frac{b}{a+b+c}$$

Write in terms of the givens:

$$\theta_1 = \frac{2N_1+N_2+N_5}{2n} = \frac{1}{5}, \theta_2 = \frac{2N_2+N_4+N_6}{2n} = \frac{1}{2}$$

where $n = \sum N_i$



Decision Rule:
$\delta = \{H_1 : T \geq c, H_2 : T < c\}$
Find c values from chi-square dist. with r - s - 1 d.o.f.
Area above c = $\alpha \rightarrow c = 7.815$

Problem 5:
There are 4 blood types (O, A, B, AB)
There are 2 Rhesus factors (+, -)
Test for independence:

	O	A	B	AB	
+	82	89	54	19	244
-	13	27	7	9	56
	95	116	61	28	300

$$T = \frac{(82 - \frac{244(95)}{300})^2}{\frac{244(95)}{300}} + ...$$

Find the T statistic for all 8 cells.
$\sim \chi^2_{(a-1)(b-1)} = \chi^2_3$, and the test is same as before.

** End of Lecture 36